南方报业集团从“相加”迈向“相融”的有益尝试 技术创新拓展媒体经营 2B 市场

——媒体大数据应用实验室牵头单位负责人接受记者采访

文 | 本刊编辑部



主持人：下面先请何崑博士介绍一下媒体大数据应用实验室，大家再分别提问。

何崑：因为国家新闻出版广电总局专门发布了国家科技标准化重点实验室，我们和百分点公司联合申报了一个。然后我们应该是整个报业为一家，有出版的，有广电的。但是在报业里面是一家。这个实验室的资源还是比较稀缺的。我们这个实验室也希望在媒体大数据应用这个方向上做一些探索，对媒体的转型起到一些推动作用。

主持人：我们的专访分为两个环节，一个是大家刚才提交的采访提纲，然后针对人工智能对行业发展先提问。

记者：您好我是中国经济周刊的记者，我还是想接着您的介绍讲一下，您介绍的言简意赅。我们目前这个联合实验室首期的项目都在做哪些事情，希望达到什么样的目的？

何崑：因为媒体大数据应用的方向还是比较多，从写作到传播效果到用户画像等等会有非常多的领域。然后我们首期可能会在用户大数据这块做一些切入，因为我们过去的传媒都是读者，但是读者不是用户。我们还是想把我们用户的一些信息，就是用户的数据能够先都收集好，然后对他们做分析，做画像。我们得了解我们现在到底是哪些用户在用我们的产品服务。然后哪些是我们应该重点去发展的方向，在哪些地方有很多用户需求，我们在不断的去提供新的服务，这个可能是我们当前需要迫切去做的，所以这块我们先去做了，同时也会做其他的事，我想先说这个。

记者：那目前利用了大数据画像之后有什么您觉得比较有意思的，或者说有价值的发现吗？

何崑：因为是两个，还有一个人工智能对媒体的理想，我们自身在我们的新媒体产品，包括网站、APP 本身都有用户数据的这些收集和整理，我们想做的更多是在整个集团的层面和范围，因为你每一个太小了。对于我们现在来说，（贸易）集团、传媒集团某个产品实际上用户数量很小，我们是把整个集团的用户数据都收集起来做一个整体的分析画像。因为那些 APP 什么网站类行为数据都已经有了，我们就做一个整体的，然后对我们整体做一个规划。我们现在是这样的，还没有做到最后已经能够落地的时候，这是我们今年要做的事情。

但是我们刚刚拿牌不久，我们在做这方面的工作，然后在这方面我们可能也会尝试着能够制定一些行业标准。然后我们看看在传媒行业里面能不能形成好的经验推广或复制到全国的同行，让他们在转型的时候也能利用这些经验。我们有一个比较好的设想，但是还没有最后到一个落地的环节，这个可能需要时间。因为我们有积累，因为包括百分点的积累，他们有一些在过去的个性推荐方面都是走在全国比较前的，我们现在利用他的一些积累和我们自身的一些数据一起来做一些探索。

记者：您这方面的研究和今日头条算法的推送有什么区别？还是说基本机制差不多呢？

何崑：其实我们做了很多其他的东西。我不想简单的用我们的用户大数据去比较，因为他这个肯定要从整个人工智能的角度，整个大数据应用角度来做。他一开始就是用算法，用模型，他们是非常重视的。我觉得他做的还是非常好的，从现在来看市场上还是非常成功的。他们做的不够好的地方可能就是他的，我们通俗一点来讲，同行叫调性，这个调性可能不一定特别好，因为这里面体现了一些社会责任的东西。

然后我觉得这里面其实也能用算法和模型来解决。我

前段时间在腾讯公司考察的时候，他们就把每篇文章再增加一些标签，他自身其实已经有很多标签来推荐推送了，然后腾讯的一个做法就是再增加一个调性的标签。调性标签以后等于开始是人工打，到了现在为止，他已经通过机器学习，70% 以上的都不由人工打了。这是一个很可怕的事情，连调性这个标签，将来都有可能通过算法、机器学习来实现。

所以我觉得今日头条如果把这个标签再打上去，其实也是很容易的事情，他也可以做这个。因为这些都是技术活，大家到了今天为止都有很优秀的技术人员，都没有问题，都能做。所以我们这块，如果你说从这个角度来讲，我们肯定也希望我们的联合实验室也能够在这方面作为我们自身产品也能够实验这样一些需要什么标签就用什么标签，什么推荐就用什么推荐。因为这个用户大数据，现在不想说的太广，因为我们其实像这些都会作为我们实验室未来的方向。用户大数据、精准推荐、个性化推荐、千人千面，还有包括内容、传播都会做。我刚刚说的只是一个开始，但是我可能不限于这块。我们的内容我想还是会比较多的。看看高总有什么补充，因为高总也是我们重要的合作伙伴。

记者：现在从媒体的角度来看的话，如果实现这种内容或者说价值的变现是媒体比较关注的一个事情。因为好多现在媒体体制改革，人工智能或者是算法这种精准推送时间之后，对这边有没有什么帮助呢？

何昆：这个是这样的，你要说内容价值变现，我们不能只是说精准推荐就实现这个价值，这个还是很有限的。我想内容价值变现可能是我们面临的一个共性问题。当然我们这块也有一定的探索。2C 可以作为一种变现，2B 可以作为一种变现。2C 的变现就拿不管是网站也好，（SP）也好，这些运营只要到了一定的流量，或者你甚至不用太多的流量，你只要你的用户群是精准的，是一个垂直领域的。你就有可能变现，你就有可能带来一些广告，包括你的信息流广告。其实我认为变现是一个比较容易的事情，不是说一个比较难的事情。

我们作为媒体自身，你要把你这个产品打造好。起码告诉你的广告商，你说我的用户是谁，我的用户群都在我这儿，那你就很容易变现，这根本就不是一个事。然后我也不是说像今日头条一样很通用的什么人都有。也许我只是广东的，或者我只是财经的，我只是汽车的这都没问题，甚至说我只是一个很小众的人群，我就是某一个地区的，我觉得你都可以变现。问题是你得有这些数据，你得有这些用户的数据告诉他们。所以说为什么我们也想从用户大数据着手的原因也是我们变现的一种方式。变现你还可以从 2B 变现，2B 的变现就是说我们现在媒体自身积累了非常多的内容，但是内容很难变现的原因就是你光这些内容不够，我们现在变现的一种方式，比如说以一种舆情为例，我们也做舆情。

在广东市场用了三年的时间签约金额达到一个亿，这些都是我们的 130 多个客户，这是 2B 的市场。然后这个我们是在我们自身那么多媒体内容的基础之上，我们还会从互联网搜索。我们现在 300 万的还有很多信息要分类整理加工，然后要服务于给我们的客户提供你要在这个基础额内容层面之上的，类似于舆情、智库这样的一些信息。我觉得你要到了这个层面的话，其实你每一个客户的产值就会比较大。我们做的最好的，我们现在三年签约一个亿，那还只是算上南方舆情的产品。其实基于这个舆情的产品上，比如说我们佛山记者站，我们佛山记者站就以这个思路去做，现在给佛山当地做智库的服务，给当地的党政企业提供智库的服务，做策划，做智库的服务。

他们去年做了五千万，据说今年的单子接不过来，就是说人手不够做不过来。我记得好像是能做到一个亿，就是一个佛山记者站。他是我们转型最好的，做的我们地方站做的最好的。我觉得这个思想就是说，你不要只是纯粹简单的那些文章，那些是不够的，你要作某一个领域。或者是就比如说党政客户关心的一些话题，你还甚至要提前预警，还要给他一些舆情研判，舆情应对，你要给他全程的服务，这个对他们来说很有价值，这是媒体最大的优势。

跑这些客户的记者知道客户的需求是什么，但其他的人不一定了解，我们了解。那你这个就是你了解需求，然后你还能够提供服务，然后不断的改变服务，用你的大数据，用你的这些分析产品提供。我觉得对他们来说非常有帮助，所以这一块就是我们目前也做了一些尝试，目前这个效果还挺好，主要是这个方向，我觉得这个路径是很对的。

另外你说变现其实还有很多的方式，比如说我们现在也积累了非常多的资源，每个媒体都有很多的历史资源库，就是媒体资源库、媒资库。媒资库现在变现比较难的一个原因，这个库里面都是非结构化的数据，你必须把这些非结构化的数据，通过人工智能的技术，通过大数据的技术提取出一些结构化的信息，然后还要做一些关联。然后慢慢的你后续就能够提供一些有价值的产品，现在只是一个单纯的搜索不够，你必须要给他提出一些结构化的信息才能做后续的一些产品。这就是为什么说现在媒资库比较难变现，但实际方向上我们已经看到了。所以我们接下来可能会做智能媒资库的项目，也是为我们后续长久的变现做准备。

记者：我们这个人工智能在大数据主要是影响媒介行业吗？他对媒介行业哪些方面有影响呢？您刚刚说更多的是用户画像，用户确立分析，其他方向有没有影响？

何昆：人工智能这个方向其实太广了，我为什么特别看好人工智能对媒体的改变和影响呢，人工智能技术一定会对媒体发生巨大的改变和影响，因为人工智能技术，我们现在说的比较多的，一个是云计算，比如说图像技术、视频技术，还有就是自然语言理解技术，这些都是跟媒体息息相关的。你说自然语言理解的技术，我们如果说从纸媒来讲，我们其实就是干的自然语言处理，这个属于人工智能非常核心的几大块之一。

这个技术里面，包括对他的分析，我刚刚提了结构化的分析，这都是自然语言处理的技术才能实现的。所以这里面我觉得光这一块都是对于媒体价值非常大的，你能提取出结构化信息才能做后续的一些内容变现。包括我们做舆情需要一些情感面的分析，这也是自然语言处理技术。你还要做一些互联网稿件分类，做一些聚类，这都是自然语言处理的技术。

所以说你有这些技术的积累，其实你后面的产品才成为

可能。之所以我们传统媒体很难变现，很难做出更多的东西，就是你没有这个技术，你甚至不知道怎么去应对这个技术，这就是过去为什么难。所以我觉得现在对于我个人来讲，我觉得我看的很清楚，未来技术和媒体的结合是非常多的。

还包括语音技术，还有现在的图像视频技术，我们现在比如说给你一个视频画面，包括比如说电视台的新闻也好，舆论也好，他可以电脑自动告诉你这是什么类别的节目，给你提取出语音的文字，甚至把画面人像给你识别出来，这是什么人，这是在电视上马上体现出这个人穿的是什么衣服，你在什么商场能买到。所以这里面未来的想象空间是非常非常广阔的，就看我们自己，你有什么商业模式，你有什么创新的产品了。我特别看好人工智能，人工智能在各个行业的应用，我觉得所有的人工智能技术基本上都能运用到媒体。所以我对这块非常非常看好。

高体伟：我稍微补充一点，我刚才讲了方方面面，如果从百分点的角度上来讲的话，恐怕我们现在关注的是两点，第一点是智能化内容的素材获取，然后能够快速的形成传播内容，这是我们发力的第一个点。第二个点，我们主要发力在智能化的传播，就是希望在恰当的时间，恰当的地点给人推恰当的内容。然后去解决信息过载还是内容过载，就是这个意思，要解决这个问题。或者希望你打开每个网站看到的内容恐怕差不多。我们现在在这两块是百分点在人工智能加媒体的一个发力点。

记者：现在咱们这个技术发展这么快，咱们也看到一些传统媒体人出现离职情况。包括我注意到我的一些弟弟妹妹，他们在报考高考专业的时候会问我，到底要不要报新闻类的专业，他们自己也会考虑包括翻译、新闻、编辑出版这些专业，可能以后都会被人工智能所取代，或者更大程度上让他们来去发挥作用。我觉得以后咱们相关专业的人才会不会有一个减少的趋势，然后咱们以后会不会还需要这些人才，然后这些人才在工作中如何跟人工智能来协同合作？

何崑：现在教育肯定跟不上了，你还是这种新闻传媒学院这种教育方法肯定全部都淘汰了。还不如文科的人去读一个计算机学院都比你在传播学院有价值。但是媒体人还是非常需要的，将来一定会更需要。但是这个对于他人的素质和要求就会有一些变化，当然他不需要了解这么深的人工智能技术，因为人工智能技术会给他提供很好的工具，他会用工具就行。

比如说他的思考，他的领域知识积累，他在某一个领域的人脉是非常非常重要的，这个不是完全靠人工智能或者是电脑能取代的。但是人工智能的技术如果应用到媒体的话，确实能够替代非常多的东西。你的什么专家库、资料库、基本信息库、关联的稿件，历史的信息，背景信息，基本上电脑里都有，甚至都可以帮你自动生成稿件。你如果只是一些普通的采访获得一点信息的话，你确确实实没有什么太大的价值。

我觉得做深度报道的肯定需要，做调查报道的都会非常需要。人工智能技术的出现，对他们来说会如虎添翼，他们会做的更精彩更出色。所以说今后的媒体人，我觉得都是非常优秀的，而且越来越宝贵，绝对不可能完全替代。

主持人：谢谢何博士演讲，我们也特别深入了了解了，人工智能和大数据技术对媒体行业的发展所带来的启示。现在是最后一个问题。或者高总可以介绍一下百分点和南方报业集团的合作和联合实验室未来的规划。

高体伟：百分点公司与南方报业从业务开展到联合一起申报国家出版业与科技的标准化实验室时，这个决定做出以后我们就开展很多思考，最重要的思考就是作为南方报业来讲，他是媒体行业里面的一个标杆。作为百分点来讲，我们是大数据里面的一个领头羊。我们想做一件事情如何能让两端结合。我当时期望的是有一个质变的过程，所以我们当时做了很多探讨。大数据如何跟媒体结合，因为在那个时候，感觉人工智能还没有现在这么火，所以我们当时重点考虑的是大数据加传媒。当时我们规划了应该是一共研讨差不多四个方向，一个方向就是如何做新闻数据的可视化，这是我们当时研究的方向。另外一个就是媒体内容的传播决策，这个也是一个方向，还有一个是传播效果的评估。最后的方向实际上刚刚何博士已经介绍了是用户的大数据，前两天我们在探讨，我们出来一个用户大数据以后，我们如何能够把这个用户大数据的体系标准化、体系化、规范化，然后从而能够上升到国家传媒业的层次进行相应的推广和指导，让整个传媒行业如何贴近用户、了解用户，能够让他们知道他们的内容，他们的价值点，以及进而延伸到他们未来价值的变现应该怎么做，这是我们对这个实验室采用的渠道。当时的整个规划应该是三年，我觉得用户大数据是本身南方报业自身的一个入口，从百分点来讲，我们现在已经开始在内容影响传播方面逐步做了。

记者：我们这块具备的优势是哪些？因为我看比如很多像国外的也在做用户画像，包括网站用户的一些数据方面的分析整理，咱们这边自身的优势体现在哪里？

高体伟：本身百分点从开始起家的时候，我们一直就在研究消费者的行为，以及消费者营销的建模。所以我们当时的第一款产品就是何博士刚刚说到的个性化推荐，个性化推荐就是我们通过四个维度来研究一个人进电子网站、电商或者内容网站的一个途径。比如说我们每个人进一个商场，我从大门进去，进去以后我是先到一楼的化妆品呢？还是到服装，再到四楼的3C家电呢？我到了一个特定场景时，即使到了四楼的电子产品，我慢慢走到3C的，慢慢走到洗衣机类的，我们往往会根据人的行为来判断在什么时间给你推荐什么样的内容或者产品。所以在这个过程里，实际上我们有很多很多的积累，画像是其中一部分，来研究这个场景，当什么场景产生的时候，这个人会做一个什么样的动作，然后才会基于我过去的历史数据。比如说我们南方报业，过去这个人在南方报业的行为，这些都会让他产生他当前的交易额。

记者：就是说你认为他这个用户的行为，其实与独特的行为习惯都是有相像地方的？

高体伟：从技术体系上是相通的，但是从人的行为上是不一样的。所以我们才要结合，来做出一个过去我们在电商所积累的画像体系，如果能够成为媒体行业标准的画像体系，这也是我们实验室要研究的一个重要工作。媒